EXHIBIT A

AVC-164

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANIZATION INTERNATIONALE DE NORMALISATION

ISO-IEC/JTC1/SC29/WG11

CODED REPRESENTATION OF PICTURE
AND AUDIO INFORMATION

ISO-IEC/JTC1/SC29/WG11
MPEG 91/ 228
1991, November

Source : Toshinori Odaka (Toshiba)
Hideyuki Ueno (Toshiba)
Hiroshi Suu (Toshiba)
Title : Proposal Package

This document includes a full proposal description for the second phase of MPEG.

CONTENTS

MS-MOTO_752_0001230293

MS-MOTO_1823_00000720813

MS-MOTO_752_0001230294

MS-MOTO_1823_00000720814

1. Introduction

The generic coding method at a bit rate of up to about 10 Mbps proposed in this document is

- a coding method aiming to improve picture quality by an MC + DCT based coding algorithm considering forward compatibility to MPEG1 and H.261 (backward compatibility is not considered) and

- a coding method which deals with the input CCIR Rec. 601/525 format as an interlaced picture format without any preprocessing and

- a coding method which uses bi-directional prediction and interpolative prediction like MPEG1.

The features of the proposed method are the following four points.

1) This method contains adaptive inter field / inter frame prediction for motion compensation, which consists of a field interpolation mode and a frame interpolation mode considering an interlaced structure.

In this method, the motion vector estimation process consists of two stages. In the second stage, some candidates for the prediction signal are calculated through a spatio-temporal filter corresponding to an adaptive prediction, compared with each other and evaluated.

2) This method adopts a special rate control method. Rate control is realized hierarchically in this method at GOP, Picture and Slice layers combining feedforward control and feedback control to cope with changes in the coded signal characteristics owing to three picture types (I-picture, P-picture and B-picture).

3) This method uses quantization characteristics in which the weight changes according to a quantizer scale.

4) This method switches the scanning order of the DCT coefficients adaptively and uses an optimized VLC for the scanning order corresponding to the interlaced format.

2. Coding algorithm

2.1 Picture format

This coding method processes a moving picture defined in CCIR Rec. 601/525 without any preprocessing as an input picture. The size of each field is shown in Fig. 2-1 and it is an interlaced format (60 fields/sec).

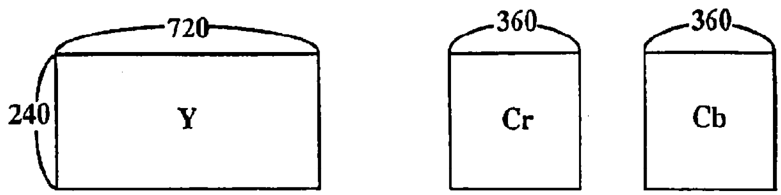


Fig. 2-1 Picture format

2.2 Structure of coded object

2.2.1 Layers

The unit of the coding operation is defined as a hierarchical structure as in MPEG1. The hierarchy is constructed from Blocks, Macroblocks, Slices, Pictures, Group of pictures (GOP), and Sequence from the bottom layer. A block contains 8 × 8 pixels and DCT is processed to this unit. A Macroblock consists of four blocks, i.e. two Y blocks together with corresponding Cr block and Cb block, and motion compensation and coding mode selection is executed to this unit. The relationship between each layer from Block to Picture is shown in Fig. 2-2.

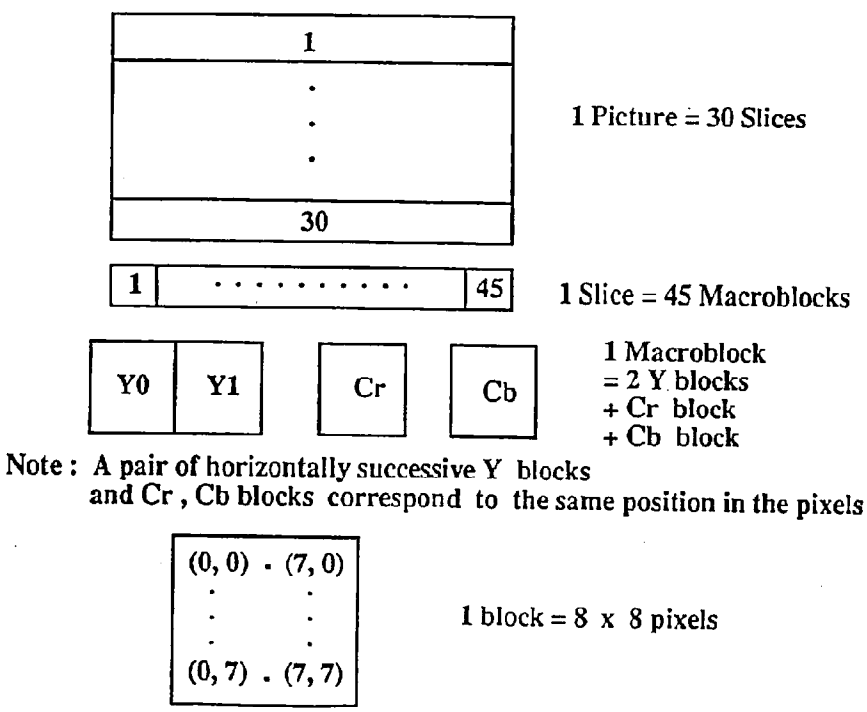


Fig. 2-2 Relationship between each layer

2.2.2 Structure of GOP

Picture is roughly categorized into three types (I-picture, P-picture, and B-picture) as in MPEG1 according to the allowed prediction mode at the Macroblock. The following four types are prepared as a predictive coding mode.

1) Intra field coding
2) Forward predictive coding
3) Backward predictive coding
4) Bi-directional interpolative predictive coding

The correspondence between each picture type and the allowed prediction mode is shown in Table 2-1.

Picture type	Allowed predictive coding mode
I-picture	Intra field coding
P-picture	Intra field coding Forward predictive coding
B-picture	Intra field coding Forward predictive coding Backward predictive coding Bi-directional interpolative predictive coding

Table 2-1 Picture type and predictive coding mode

In this proposal, an interlaced format is adopted as the coding picture format, and a different prediction method is applied even to the same type of picture according to its position in GOB. So, P-picture and B-picture is categorized further in detail into P0, P1, P2 and B0, B1, B2, B3, respectively. An I-picture always exists in an even field. The structure of GOP is shown in Fig. 2-3. As shown in the figure, GOP is defined as a set of pictures from the B0-picture four pictures before an I-picture to the P2-picture five pictures before the next I-picture.

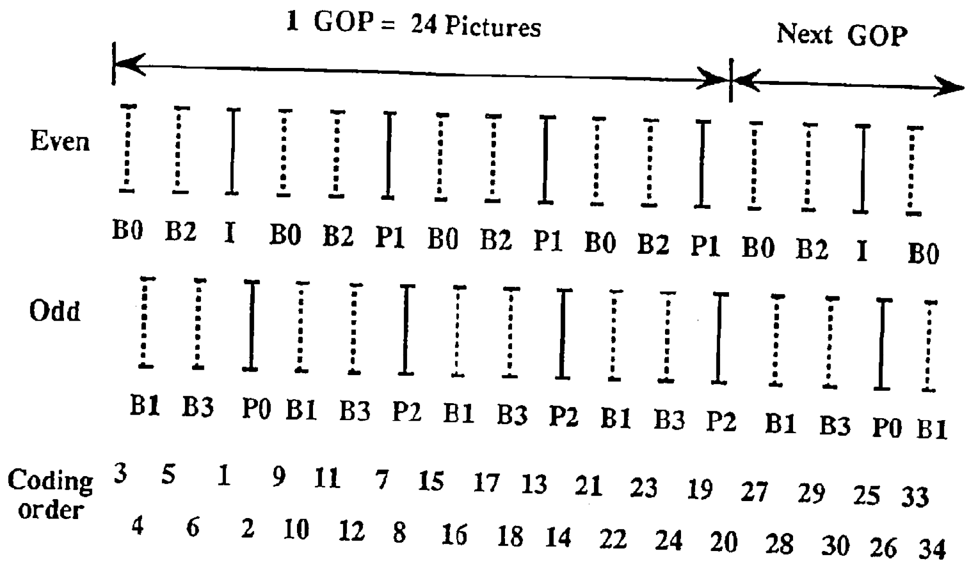


Fig. 2-3 Structure of GOP

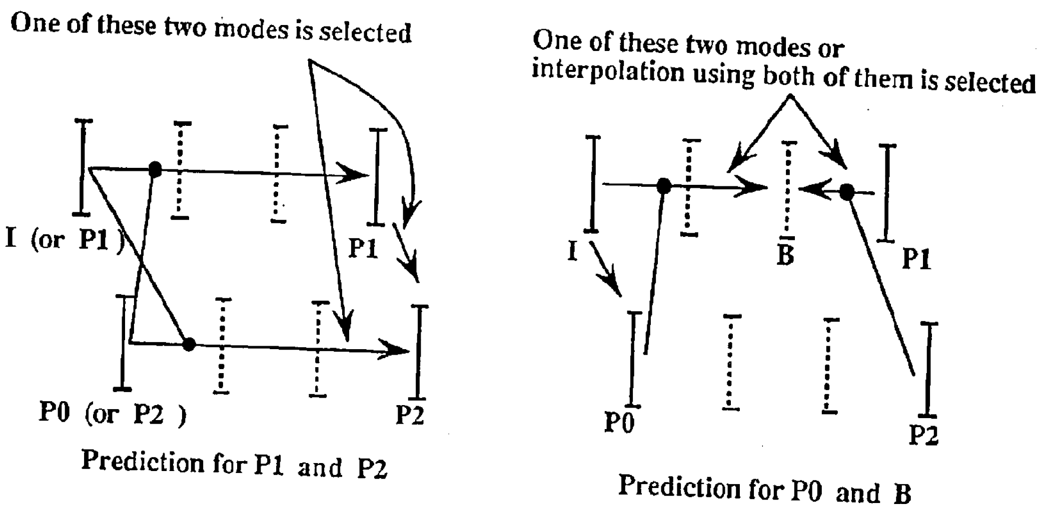


Fig. 2-4 Prediction for each picture

Table 2-2 and Fig. 2-4 show how and from which picture each picture is predicted. There are two types of prediction in forward prediction coding. One is inter field prediction from only an even field, and the other is adaptive inter field / inter frame prediction from a set of even and odd field. In Fig. 2-4, $\longrightarrow$ denotes inter field prediction and $\nearrow\!\!\rightarrow$ denotes adaptive inter field / inter frame prediction.

The coding order of each picture in GOP is also shown in Fig. 2-3.

Picture type	Prediction method and reference picture
I-picture	Without prediction
P0-picture	Inter field forward prediction from I
P1-picture	Adaptive inter field / inter frame prediction in forward direction from past P1 and P2 (or I and P0)
P2-picture	Selection from Adaptive inter field / inter frame prediction in forward direction from past P1 and P2 (or I and P0) and Inter field forward prediction from P1
B0-picture to B3-picture	Selection from Adaptive inter field / inter frame prediction in forward direction from past P1 and P2 (or I and P0) and Adaptive inter field / inter frame prediction in backward direction from P1 and P2 (or I and P0) in the future and Bi-directional interpolative prediction from past P1 and P2 (or I and P0) and P1 and P2 (or I and P0) in the future

Table 2-2 Prediction method and reference picture for each picture type

2.3 Block diagram and outline of coding process

The proposed algorithm is basically based on motion compensation and DCT coding. Figure 2-5 shows the encoder block diagram of the proposed method. Some input fields are stored in individual field memories. These fields are sent to a motion vector estimator 1, and here the motion vector at the first stage is estimated in one pixel precision using pictures before coding by telescopic search (cf. 2.4). Following this, the estimated vector is refined by fully searching the range of ± 1 around the estimation in the first stage as preprocessing for the second stage, using locally decoded pictures in field memories in the local decoding loop. Next, the second stage is as follows. In this stage, the motion vector in half pixel precision is estimated at motion vector estimator 2, a filter for adaptive prediction, and field memories to determine the prediction signal. Prediction error is processed through the DCT, quantizer, adaptive scanner, and two-dimensional variable length coder, and sent to the multiplexer. On the other hand, the quantized data is used for local decoding through the dequantizer and inverse DCT. Only locally decoded signals for the I-picture and P-picture are stored in the field memory. The number of field memories necessary in the local decoding loop is four for preparing adaptive prediction signals. The coefficient data is multiplexed at the multiplexer to side informations such as the motion vector, Macroblock type, quantizer scale, etc. The multiplexed information is output through a buffer. The coding controller determines the quantizer scale value according to the amount of buffer contents and the activity of the input picture pre-examined before coding (for I-picture) or the activity of the adaptive prediction error signal obtained in the coding process (for P and B-pictures).

Fig. 2-6 shows a block diagram of the processing element of the adaptive inter field / inter frame predictor. Reference data for adaptive prediction are read out from the decoded field memory. The accessing position of the reference is pointed by the corresponding motion vector from the motion vector estimator 1. The forward (backward) prediction signal is generated by operating a spatial filter or spatio-temporal filter to one or two of these references or by selecting one of the references itself. An actually used prediction signal is that selected from the forward prediction signal, backward prediction signal, and their average.

Fig. 2-7 shows the decoder block diagram of the proposed method. The DCT coefficient signal is decoded by the variable length decoder. After this, the same decoding process as that of the local decoder in the encoder follows. The motion vector transmitted as a difference vector is reformed to a normal vector after being variable length decoded.

The output of the pictures for display is switched between the I and P-pictures from the field memory and the B-picture directly from the decoding process, because the decoding order and displaying order is different, and also because it is not necessary to store a B-picture which is not used as a reference for prediction.

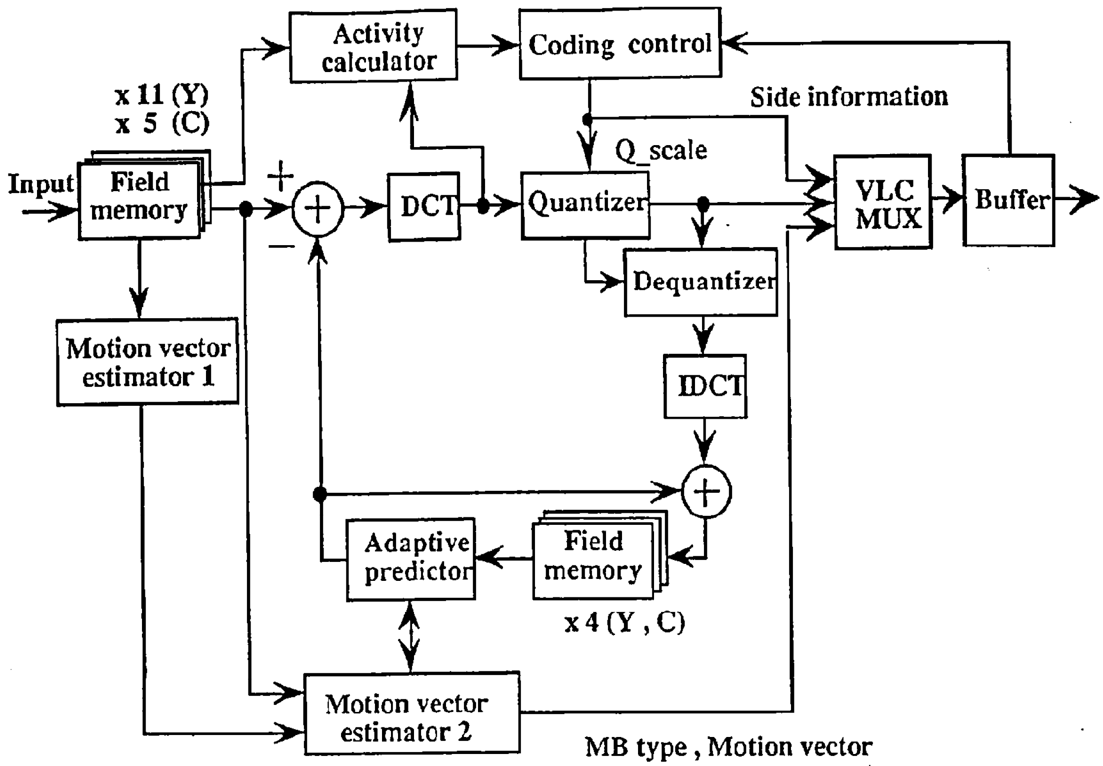


Fig. 2-5 Encoder block diagram

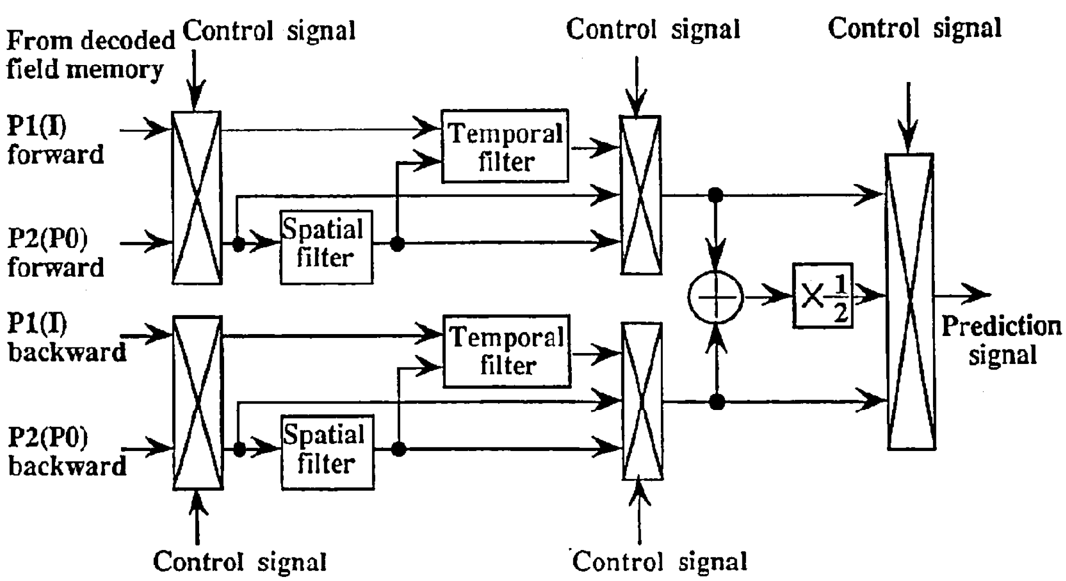


Fig. 2-6 Adaptive inter field / inter frame prediction processing element

MS-MOTO_752_0001230301

MS-MOTO_1823_00000720821

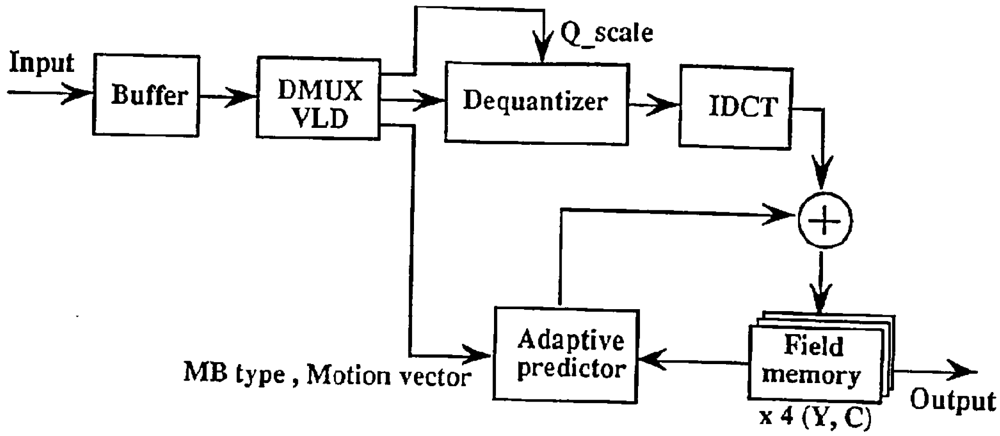


Fig. 2-7 Decoder block diagram

2.4 Motion compensation and search for motion vector

Motion compensation is applied to each Macroblock. The motion estimation process consists of two stages in this proposal i.e. the usual motion estimation in one pixel precision by block matching (1st stage) and motion estimation by searching all candidates for the prediction pixel adaptively interpolated in half-pixel precision around the first estimated motion vector (2nd stage). A more precise way for adaptive interpolation is explained in the next section.

The first motion estimation stage is a process to search an optimal motion vector in one pixel precision for each field comparing the input pictures. Telescopic search (cf. MPEG1 SM3 specification) is applied to a set of fields whose distance is more than two fields. Note that there are plural search paths in case of interlaced pictures. In this proposal, the search path is determined according to the following rules.

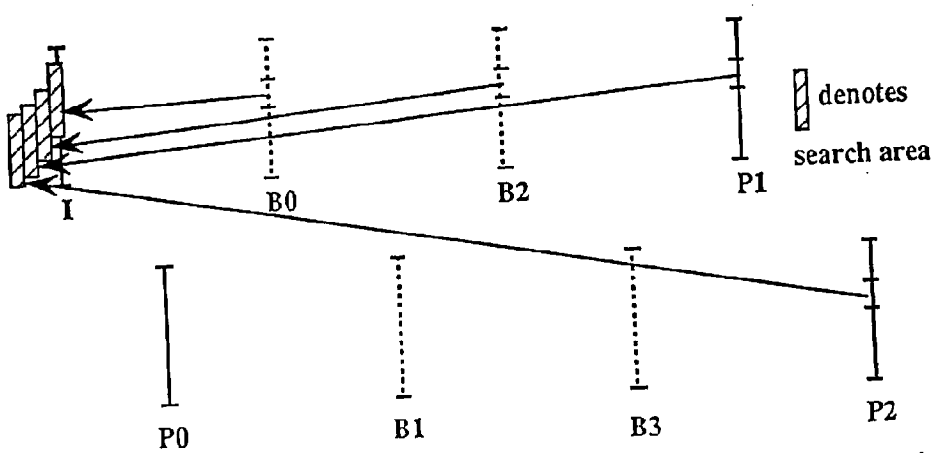
1) Search between fields with the same interlacing phase, even or odd
   Use only fields with the same interlacing phase.
2) Search between fields with a different interlacing phase
   Use as little fields with a different interlacing phase as possible. It is necessary to contain in the whole search path the search to a field with a different interlacing phase at least once. In this proposal, the searching order is selected so as to search the fields with a different interlacing phase as the last search as shown in Fig. 2-8.

MS-MOTO_752_0001230302
MS-MOTO_1823_00000720822

Telescopic search is applied to forward and backward directions independently. The size of the search area at each search step is ±15 for the horizontal direction and ±7 for the vertical direction.

The second stage motion estimation is executed using pictures in the field memories in the local decoding loop as reference. Before the main process, the motion vector estimated at the first stage is refined by fully searching the range of ±1 around the obtained motion vector as a preprocess. The main process of the second stage is executed by selecting an optimal prediction signal from the candidates of the prediction signal calculated around the refined vector in half-pixel precision by the method explained in the next section, evaluating and comparing each power of prediction error.

Motion estimation is not executed for the chrominance signal, but motion compensation is applied using the motion vector calculated from the motion vector used for the corresponding luminance signal.



ex. When motion vector from I to P is searched, the searching order is
I ⟶ B0 ⟶ B2 ⟶ P1 ⟶ P2

Fig. 2-8 Searching order in " telescopic search "

2.5 Adaptive inter field / inter frame prediction

The main process of the second stage motion estimation is a search in half pixel precision around the optimal vector obtained in the preprocess. Figure 2-9 shows how the process is carried out. This adaptive inter field / inter frame prediction is executed for P1, P2, and B-pictures, and uses a pair of even and odd fields as reference fields. In Fig. 2-9, suppose that the optimal vectors for the individual fields (field #1 and #2) are obtained as v1 and v2 in the first stage (denoted by ● in the figure). Search is executed around these two points in the second stage. There are two modes prepared for generating prediction signal candidates in the second stage. One is the field interpolation mode and the other is frame interpolation mode.

Candidates are generated only by a spatial filter in the field interpolation mode. For example in the figure, intra field interpolated pixel values p1 to p3 are calculated by the following formula.

$$P1 = \frac{1}{2} (P0 + P4)$$
$$P2 = \frac{1}{4} (P0 + P4 + P5 + P6)$$
$$P3 = \frac{1}{2} (P0 + P6)$$

9 candidates are generated in this mode for even and odd fields. So, a total of 18 candidates are available.

Candidates are generated in the frame interpolation mode by operating a spatio-temporal filter to the pixels motion compensated in 1 pixel precision at each field. For the example in the figure, inter frame interpolated pixel values p1' to p3' are calculated by the following formula.

$$P1' = \frac{1}{2} \times P7 + \frac{1}{4} (P1 + P4)$$
$$P2' = \frac{1}{2} \times P7 + \frac{1}{8} (P0 + P4 + P5 + P6)$$
$$P3' = \frac{1}{2} \times P7 + \frac{1}{4} (P0 + P6)$$

We call such a field a base field that provides the pixel value in half-pixel precision and corresponding vector a base vector. There are also 9 candidates in this mode generated for even and odd field, but the candidates for position 0 for each field give the same value. So, a total of 17 candidates are available in this mode.

All of these 35 candidates are evaluated in the second stage, compared, and a candidate which gives the least prediction error is selected as the prediction signal. However, candidates from the frame interpolation mode is not selected when the direction of the motion vector for both fields are very different (more precisely explained later). A one bit flag is transmitted to show which mode is selected. Another one bit flag is transmitted to show which field is selected in the field interpolation mode or which field is selected as a base field in the frame interpolation mode. Inter field prediction in P0 and P2-pictures is carried out by applying the field interpolation mode to a single field. No flag is transmitted in this case. In the B-picture case, only a field with a different interlacing phase from a field to be coded can be selected as the base field in the frame interpolation mode. No flag is transmitted to show which field is selected as the base field.

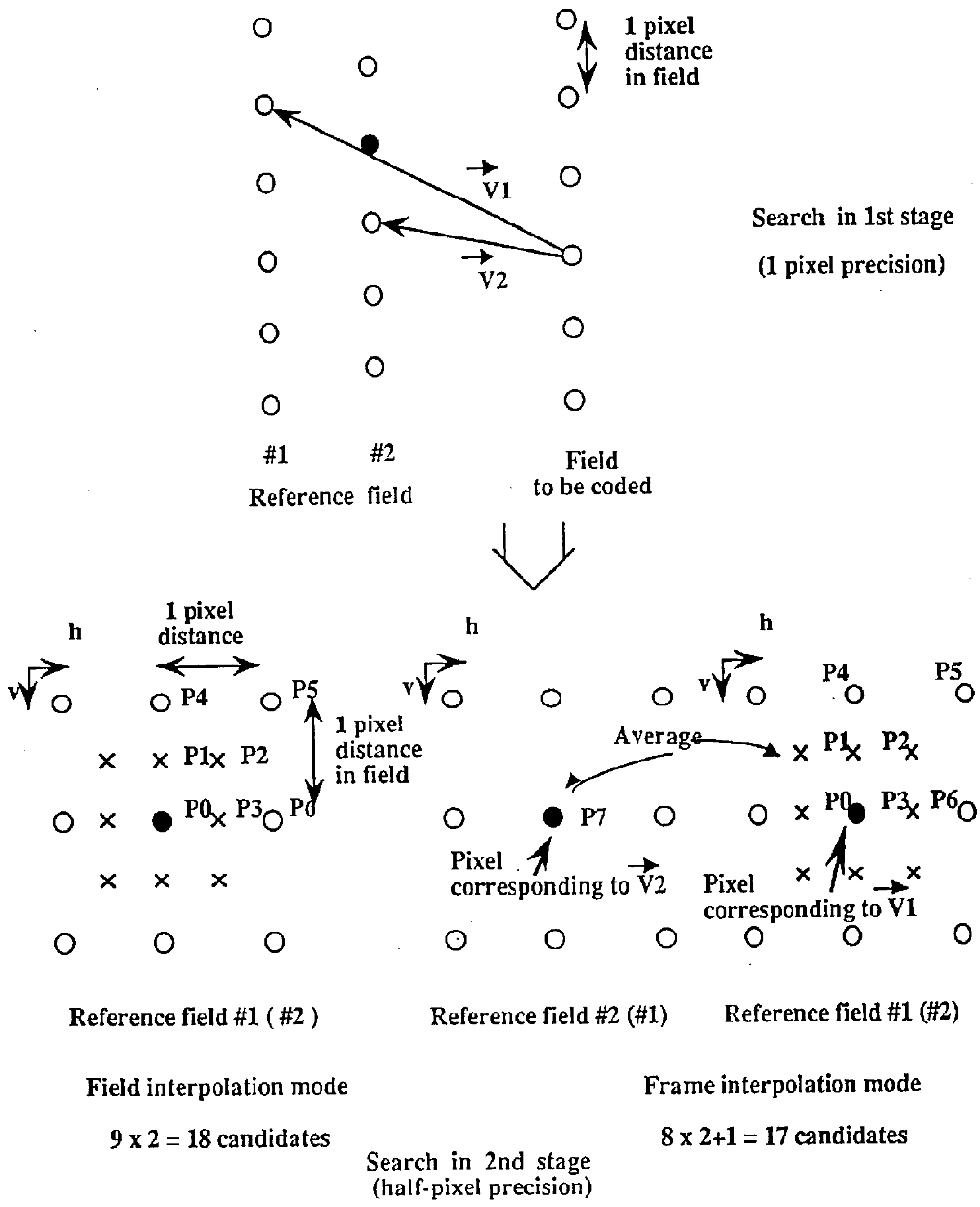


Fig. 2-9 Adaptive inter field / inter frame prediction

Motion vectors are transmitted in the following forms. First, the motion vector in the field interpolation mode or the base motion vector in the frame interpolation mode is transmitted in half-pixel precision. When the frame interpolation mode is selected, the motion vector other than a base vector follows as a difference vector to the base vector. The motion vectors to both fields likely to give a similar direction in this case. So, the rule to transmit the difference vector is determined as follows. First, the cross point of the expanded base motion vector and fields other than the base field is calculated. Next, the motion vector in one pixel precision corresponding to the nearest point to the cross point is calculated. Then, the difference between the vector other than the base vector and the calculated vector is transmitted in one pixel precision. The frame interpolation mode is inhibited when the difference vector exceeds the range of ±1 (see Fig. 2-10). By applying this rule, the amount of motion vector information can be saved without degrading the performance of prediction, because the frame interpolation mode is useful only when the direction of the two vectors is nearly the same.

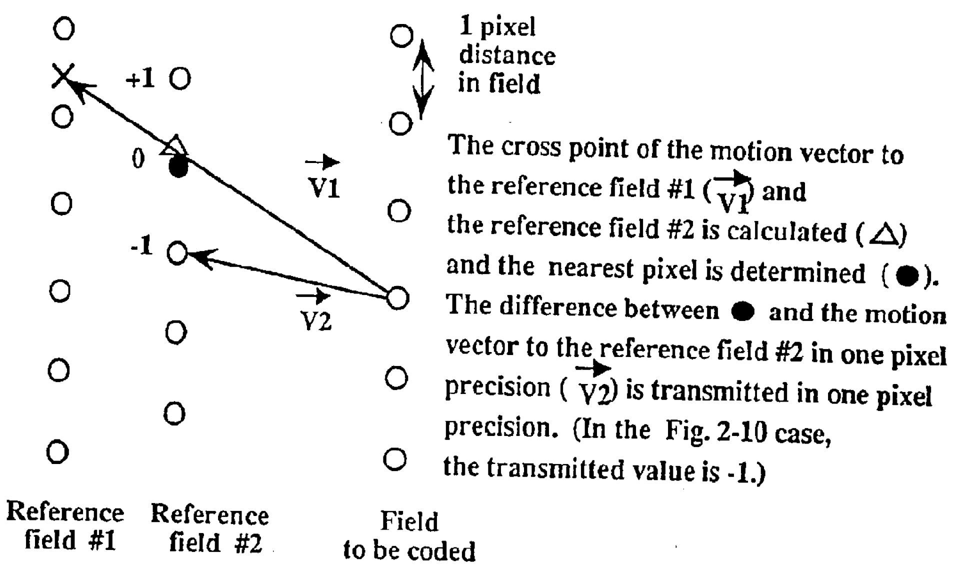


Fig. 2-10 Transmission of motion vector

Motion compensation for the chrominance signal is executed based on the motion vector obtained for its corresponding luminance signal. The luminance signal and the chrominance signal have the same number of pixels in the vertical direction but the chrominance signal has half the pixels of the luminance signal in the horizontal direction in the same Macroblock. So, when the motion vector is applied to the chrominance signal, the horizontal component of the vector is divided by 2. The division is executed so that any fractional part is rounded to the direction toward zero. This situation is the same for both the field interpolation mode and the frame interpolation mode. An example of the process is shown in Fig. 2-11. In this figure, circles drawn by a dotted line show the pixel position at which no chrominance signal exists. Suppose that the

center position of nine circles corresponds to the motion vector obtained in the first stage and x is that obtained in the second stage. Suppose also that the horizontal component of the center circle is smaller than that of the origin in this example. In the field interpolation mode, when the motion vector is obtained by dividing the motion vector for the luminance signal by 2 at the x position, the quarter pixel precision component is rounded to the direction toward zero, and the interpolated value is calculated for the position shown by Δ. This means that the prediction signal is calculated by the following formula.

$$\triangle = \frac{1}{4}(q1 + q2 + q3 + q4)$$

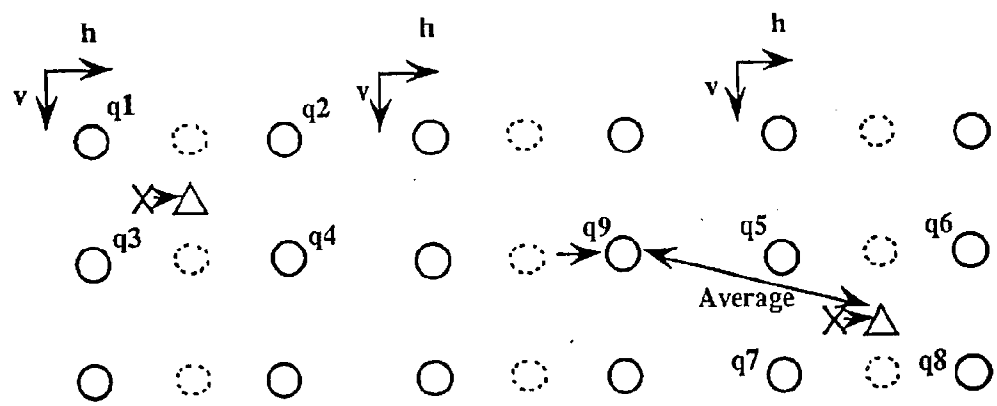
The same calculation rule for the prediction signal is applied to the reference field #1 in the frame interpolation mode. This means that the interpolated pixel value is calculated for the Δ position by the following formula.

$$\triangle = \frac{1}{4}(q5 + q6 + q7 + q8)$$

On the other hand, the interpolation pixel value in the reference field #2 becomes the pixel value of the pixel at position 9, because the half-pixel precision component of the motion vector for chrominance signal is rounded to the direction toward zero.

It is concluded in this example that the prediction signal for the chrominance signal in the frame interpolation mode is calculated by the following formula.

$$\frac{1}{2} \times q9 + \frac{1}{8}(q5 + q6 + q7 + q8)$$



Reference field #1 Reference field #2 Reference field #1

Field interpolation mode Frame interpolation mode

Fig. 2-11 Motion vector for chrominance signal

2.6 Macroblock type and type selection

2.6.1 I-picture

The variation in the Macroblock type in the I-picture is the same as that of the intra frame in MPEG1. In this simulation, Intra with modified Q is not used.

2.6.2 P-picture

The variation in the Macroblock type and rules for type selection in P0 and P1-pictures are the same as those in the predicted frame in MPEG1 SM3.
The variation of the Macroblock type in the P2-picture is the following.

- Near-field predicted, not coded
- Near-field predicted, coded (with modified Q)
- Far-frame predicted, not coded
- Far-frame predicted, coded (with modified Q)
- Intra (with modified Q)

'Near-field predicted' denotes inter field prediction using only the immediately prior P1. 'Far-frame predicted' denotes inter field / inter frame adaptive prediction using I (P1), 7 fields before and P0 (P2), 6 fields before. Selection from these is executed based on the decision tree shown in Fig. 2-12. Note that this figure is generated by replacing 'Forward' and 'Backward' to 'Near field' and 'Far frame' and removing the branch corresponding to the interpolation mode in the decision tree for the interpolated frame in MPEG1 SM3. In this simulation, modes with modified Q are not used.

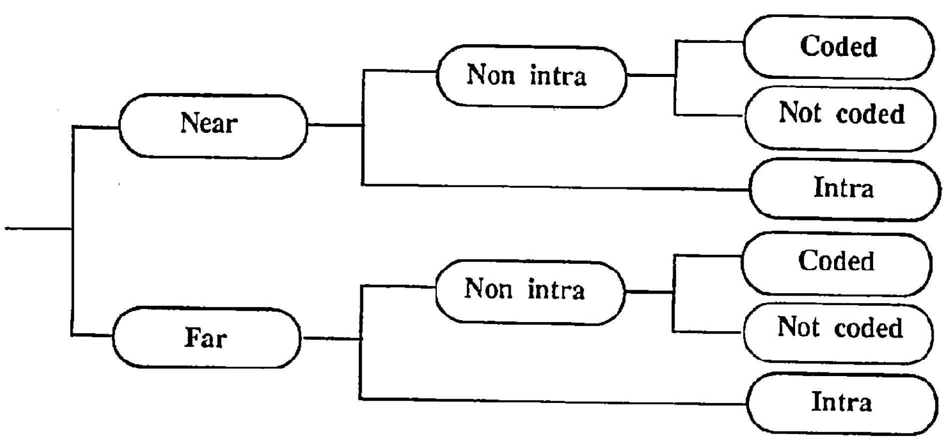


Fig. 2-12 Decision tree for P2-picture

2.6.3 B-picture

The variation in the Macroblock type and rules for type selection in the B-picture are the same as those in the interpolated frame in MPEG1 SM3.

MS-MOTO_752_0001230308
MS-MOTO_1823_00000720828

2.7 Quantization

This proposal utilizes the quantization property with which the ratios between any two of the weighting coefficients for each sequency of DCT coefficients change according to the quantizer scale value. The quantizer for each sequency is a linear quantizer without a dead zone in Intra case and with a dead zone in a Non-intra case (same as those of MPEG1 SM3).

Figure 2-13 shows the relationship between the step size and the sequency number at each quantizer scale. These properties are obtained by (number read from a fixed matrix) / 8 × (quantizer scale) + (offset). The fixed matrices are shown in Fig. 2-14.

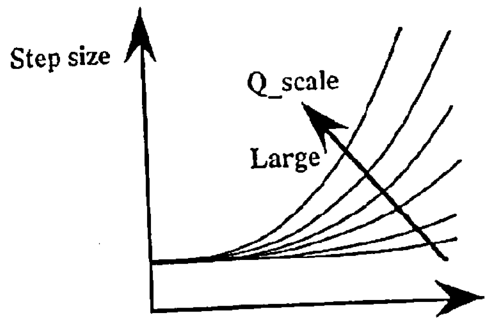


Fig. 2-13 Quantization property

By using this quantization property, the less the usable bit rate becomes and coarser the selected quantizer becomes, the lower the priority of the higher sequency compared to the lower sequency becomes. The feature of this quantization property is that the relationship between the quantizer scale and the step size at each sequency is represented by a non-linear function, and the ratio of the step size for low sequency and that for high sequency changes according to the quantization scale. The offset values are 8 for Intra and 2 for Non-intra.

0	8	11	14	18	19	21	26
8	8	14	16	19	21	26	29
11	14	18	19	21	26	26	30
14	14	18	19	21	26	29	32
14	18	19	21	24	27	32	40
18	19	21	24	27	32	40	50
18	19	21	26	30	38	48	61
19	21	27	30	38	48	61	75

Intra

8	10	12	14	16	18	20	22
9	11	13	15	17	19	21	23
10	12	14	16	18	20	22	24
11	13	15	17	19	21	23	25
12	14	16	18	20	22	24	26
13	15	17	19	21	23	25	27
14	16	18	20	22	24	26	28
15	17	19	21	23	25	27	29

Non-intra

Fig. 2-14 Weighting matrices to generate quantization properties

2.8 Adaptive scanning

This proposal utilizes adaptive switching of the scanning order of the DCT coefficients, because the distribution of the power of the DCT coefficients show different properties from that of CIF in H.261 and SIF in MPEG1 owing to the interlaced structure of the input signal. Switching is executed automatically by the condition combining Intra / Non-intra and luminance / chrominance, so no extra side information is necessary to show the scanning order. The scanning orders for the individual conditions are shown in Fig. 2-15.

0	2	6	12	20	28	34	50
1	4	11	19	27	33	35	51
3	8	17	25	31	36	49	52
5	10	18	26	32	37	48	53
7	15	23	30	39	38	47	54
9	16	24	40	45	46	55	60
13	21	29	41	44	56	59	61
14	22	42	43	57	58	62	63

Intra Y

0	1	5	6	14	15	27	28
2	4	7	13	16	26	29	42
3	8	12	17	25	30	41	43
9	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54
20	22	33	38	46	51	55	60
21	34	37	47	50	56	59	61
35	36	48	49	57	58	62	63

Intra C

0	8	16	24	32	40	48	56
1	9	17	25	33	41	49	57
2	10	18	26	34	42	50	58
3	11	19	27	35	43	51	59
4	12	20	28	36	44	52	60
5	13	21	29	37	45	53	61
6	14	22	30	38	46	54	62
7	15	23	31	39	47	55	63

Non-intra Y

0	1	2	3	4	7	15	34
5	6	8	10	14	26	35	56
9	11	13	16	25	36	43	58
12	18	17	24	30	42	50	59
19	23	27	29	41	48	52	60
20	28	31	37	44	49	54	61
21	32	38	40	46	51	55	62
22	33	39	45	47	53	57	63

Non-intra C

Fig. 2-15 Scanning order table

2.9 Variable length coding

1) Macroblock address

The same VLC table as that of MPEG1 is used.

2) Macroblock type

For I, P0, P1, and B-pictures, the same VLC table as that for the I, P and B-pictures, respectively, in MPEG1 is used. As for P2-pictures, the same VLC table as that for the B-picture in MPEG1 SM3 interpreting 'Forward' and 'Backward' as 'Near-field' and 'Far-frame' is used.

3) Motion vector

- The motion vector in the field interpolation mode or the motion vector for the base field in the frame interpolation mode (half-pixel precision in both cases) is transmitted as the difference vector to that of the motion compensated Macroblock placed left to the attended Macroblock selected according to the priority shown in Table 2-3. The same VLC table as that of MPEG1 is used to transmit the difference vector.

- In case of the frame interpolation mode, the motion vector for fields other than the base field as the difference vector to the motion vector for the base field is transmitted in a limited range of ±1 in one pixel precision by the VLC table shown in Table 7-4g.

Priority 1 When the previous Macroblock is a motion compensated Macroblock and the interlacing phase of the attended field and the selected field in the field interpolation mode or the base field in the frame interpolation mode is the same, take the difference with the vector to that field.

Priority 2 When the previous Macroblock is a motion compensated Macroblock and the attended Macroblock is predicted by the frame interpolation mode and the interlacing phase of the attended field and the base field is different, reconstruct the motion vector to the field with the same interlacing phase from the difference vector and make a difference with that vector.

Priority 3 When the previous Macroblock is a motion compensated Macroblock and the attended Macroblock is predicted by the field interpolation mode and the interlacing phase of the attended field is different from the field selected for prediction, make a difference with the vector to that field.

Priority 4 When the previous Macroblock is not a motion compensated Macroblock, apply the above 1 to 3 to one Macroblock before. Added to this, the same reset rule as that of MPEG1 is applied to the reference motion vector.

Table 2-3 Priority of selection for subtracted motion vector

4) Coded block pattern

The pattern number defined by the following formula are coded by the VLC table shown in Table 7-3.

$$\text{Pattern No.} = 8 \times y0 + 4 \times y1 + 2 \times cr + cb$$

where y0, y1, cr, and cb give the value 1 when blocks Y0, Y1, Cr, and Cb are coded, and 0 when they are not coded, respectively.

5) DCT coefficient

DCT coefficients are 2-dimensional variable length coded in the form of zero run + non-zero coefficient (= (RUN, LEVEL) : called an event). There comes an EOB (= End of block) code at the end of each block. The VLC code word for EOB is also prepared in the VLC table. There are some difference in the way transmitting between Intra Macroblock case and Non-intra Macroblock case.

- Intra Macroblock case

DC components are transmitted as a difference to that of the previous block. Rules for how to set and reset the DC predictor is the same as that of MPEG1 SM3. AC components are transmitted by the 2-dimensional VLC for Intra AC coefficients.

- Non-intra Macroblock case

Two types of 2-dimensional VLC tables are switched for the first event and the others. The VLC table applied to events other than the first event is the same as that for Intra AC coefficients.

2.10 Rate control

2.10.1 Outline

The proposed rate control method consists of the following 3 layers.

1) Bit allocation for each GOP.
2) Bit allocation for each picture and update of the allocated number of bits.
3) Rate control in each picture using a hypothetical buffer.

A value called activity is calculated and used for rate control. The activity for an I-picture is the sum of the absolute values of the pixels subtracting the over-block mean value. The activity for a P or B-picture is the sum of the absolute values of the DCT coefficients of adaptive prediction error signal. The activity for an I-picture is calculated before coding an I-picture. The activity for a P or B-picture is calculated while coding, and the activity for the most recently coded picture of the same type is used for rate control.

2.10.2 Bit allocation for GOP

Rate control is carried out so that the number of generated bits within a GOP approaches the target value C_{GOP}. The leftover bits are reallocated to the next GOP. The number of allocated bits for the i-th GOP, $R_{GOP}[i]$, is calculated as follows:

$$R_{GOP}[i] = C_{GOP} + (R_{GOP}[i - 1] - G_{GOP}[i - 1])$$
$$C_{GOP} = R \times N / 60$$

where R is the coding rate (bits/second), N is the number of picture in a GOP, $R_{GOP}[i-1]$ and $G_{GOP}[i-1]$ are the number of allocated bits and generated bits in the (i-1)-th GOP, respectively.

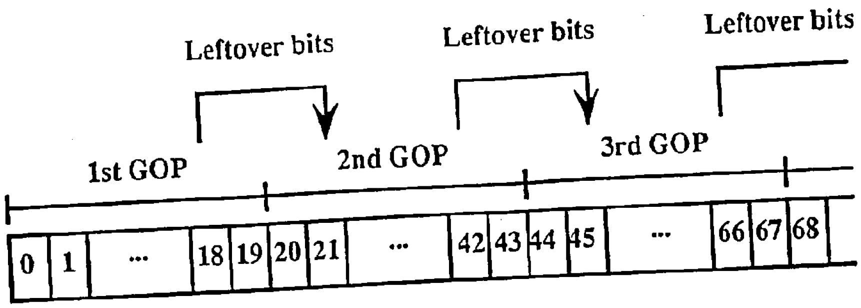


Fig. 2-16 Bit allocation for GOP

2.10.3 Bit allocation for each picture

2.10.3.1 Bit allocation for I-picture

The number of allocated bits for the first I-picture in a sequence, R_{I_0}, is calculated as follows:

$$R_{I\_0} = (4.28\times10^{-8} \times A_{I\_0} + 9.00\times10^{-2}) \times C_{GOP}$$

where A_{I_0} is the activity for the first I-picture.

The number of allocated bits for other I-pictures, R_I, is calculated as follows, so that the ratio of the average quantizer scale in P1 and P2-pictures to that in a I-picture approaches to the target ratio, K_{QI_P12}:

$D_{I\_P12}' = Q_I' - (Q_{P1}' + Q_{P2}') / 2 / K_{QI\_P12}$
if(D_{I_P12}' >= 1) {
  $R_I = G_I' \times A_I / A_I' \times Q_I' / (Q_I' - 1);$ $Q_{I,S\_first} = Q_I' - 1;$
} else if(D_{I_P12}' <= -1) {
  $R_I = G_I' \times A_I / A_I' \times Q_I' / (Q_I' + 1);$ $Q_{I,S\_first} = Q_I' + 1;$
} else
  $R_I = G_I' \times A_I / A_I';$

A_I and A_I' are the activities for the attended and the most recently coded I-picture respectively, G_I' is the number of generated bits in the most recently coded I-picture, Q_I', Q_{P1}' and Q_{P2}' are the averages of the quantizer scales in the most recently coded I, P1 and P2-pictures, respectively, and K_{QI_P12} is defined in Table 2-4. Q_{I,S_first} is the quantizer scale used in the first slice of the concerned I-picture, and it is used in rate control in the I-picture. The maximum value of R_I is limited to 40% of C_{GOP}.

2.10.3.2 Bit allocation for P0-picture

Bit allocation is not explicitly applied to a P0-picture. However, as will be mentioned later, the number of generated bits in a P0-picture is limited, so that the sum of the number generated in a P0-picture and that in the I-picture before the P0-picture does not exceed 50% of C_{GOP}.

2.10.3.3 Bit allocation for P1, P2, and B-pictures

The number of allocated bits for a P1, P2 or B-picture is determined as follows:

$$R_{P1} = (R_{GOP} - G_C) / (N_{P1} + K_{P1\_P2} \times N_{P2} + K_{P1\_B} \times N_B)$$
$$R_{P2} = K_{P1\_P2} \times R_{P1}$$
$$R_B = K_{P1\_B} \times R_{P1}$$

where G_C is the number of bits generated in pictures that are already coded in the GOP, N_{P1}, N_{P2}, and N_B are the number of P1, P2 and B-pictures that have not been coded in the GOP respectively, K_{P1_P2} and K_{P1_B} are the ratio of the number of allocated bits for the P2-picture to that for the P1 picture and the ratio of the number of allocated bits for the B-picture to that for the P1-picture, respectively. K_{P1_P2} and K_{P1_B} have initial values 0.750, 0.375, and are updated just after coding a B3-picture as follows.

K_{P1_P2} is updated by the following equation using the activities for the most recently coded P1 and P2-pictures, A_{P1}' and A_{P2}'.

$$K_{P1\_P2} = 1.85 \times A_{P2}' / A_{P1}' - 0.83$$

K_{P1_B} is updated by the following equations so that the average quantizer scale of a B-picture will be K_{QP12_B} times larger than those of P1 and P2-pictures.

$$K_{P1\_B} = G_{P1}' / G_B'$$
$$G_B' = \sum_{k=0}^{3} (1/4)^{DBk'} \times G_{Bk}' / \sum_{k=0}^{3} (1/4)^{DBk'}$$
$$D_{Bk}' = 0 \qquad k=0$$
$$= | Q_{B0}' - Q_{Bk}' | \qquad k=1,..3$$
$$Q_{Bop}' = (Q_{P1}' + Q_{P2}') / 2 \times K_{QP12\_B}$$

where G_{Bk}' is the number of generated bits in the most recently coded Bk-picture, Q_{P1}', Q_{P2}', and Q_{Bk}' are the average of the quantizer scales in P1, P2, and Bk-pictures respectively, and K_{QP12_B} is defined in Table 2-4.

Input sequence	K_{QI_P0}	K_{QI_P12}	K_{QP12_B}
"Mobile & calendar"	1.0	1.5	2.0
Other sequences	1.0	1.0	1.5

Table 2-4 Target quantizer scale ratio

2.10.4 Rate control in each picture

Rate control in each picture is carried out slice by slice, using a hypothetical buffer. The quantizer scale used in the first slice of each picture is summarized in Table 2-5.

I:	$0.9 \times A_{I\_O} / R_{I\_O}$	(First I-picture in sequence)
	Q_{I,S_first}	(Other I-Pictures)
P0:	$Q_I' \times K_{QI\_P0}$	
P1:	$Q_{P0}' \times K_{QI\_P12} / K_{QI\_P0}$	(First P1-picture in sequence)
	$Q_{P1}' \times G'_{P1} / R_{P1}$	(Other P1-pictures)
P2:	Q_{P1}'	
B0:	$(Q_{P1}' + Q_{P2}') / 2 \times K_{QP12\_B}$	
B1:	Q_{B0,S_end}	
B2:	Q_{B1,S_end}	
B3:	Q_{B2,S_end}	

Table 2-5 Initial value of quantizer scale

2.10.4.1 Rate control in I, P1, and P2-pictures

The content of the hypothetical buffer is set to zero before coding each picture. The generated bits in each slice, whose number is G[slice], are accumulated in the buffer, and the allocated bits for the slice, whose number is R[slice], are taken from the buffer. R[slice] is determined so that the allocated bits for the picture, whose number is R_{pic}, are allocated to each slice proportionally to the activity for the slice, A[slice].

$$R[slice] = R_{pic} \times A[slice] \Big/ \sum_{m=1}^{NUM\_SLICE} A[m]$$

where NUM_SLICE is the number of slices in a picture. As mentioned above, the activity for an I-picture is calculated before coding the I-picture, and the activities for other pictures are those calculated while coding. The buffer content B[slice] is referenced slice by slice. If B[slice] is larger (smaller) than both a threshold T (-T) and the buffer content at the previous slice, B[slice-1], then the quantizer scale Qs is increased (decreased) by one. Qs is limited to the range from 1 to 31. These processes are summarized as follows: